3D-EMCAD：用于体积脑肿瘤分割的高效多尺度卷积注意力解码

Iftakhar Hossain　Israt Jahan Sinthia　生物医学工程系　生物医学工程系　孟加拉国工程与技术大学（BUET）　吉大港工程与技术大学（CUET）　达卡，孟加拉国　吉大港，孟加拉国
iftakhar.bme.buet20@gmail.com　sinthiasrt@gmail.com

Md. Hadiur Rahman Khan　电气与电子工程系
孟加拉国工程与技术大学（BUET）　达卡，孟加拉国
hrkhan@eee.buet.ac.bd

arXiv:2611.29027v1 [eess.IV] 21 Nov 2026

摘要—**编码器—解码器网络广泛用于体积脑肿瘤分割，但近期趋势倾向于采用高容量解码器，这类解码器在高空间分辨率下进行三维处理时会带来高昂的计算成本。高效多尺度卷积注意力解码器（EMCAD）是一种轻量级解码器，能够以极低的成本实现出色的准确度，但目前仅在二维场景中得到验证。本文将 EMCAD 扩展至体积分割（3D-EMCAD），将所有运算转换为三维，并用组归一化替换标准归一化层，以适应体积图像训练中较小的批量大小。为单独考察解码器选择的影响，我们在 BraTS 2020 上采用相同的 3D SegResNet 编码器和训练方案，评估五种解码器：所提出的 3D-EMCAD、原始 SegResNet 解码器、Attention U-Net 解码器、UNet++ 和标准U-Net解码器（369 例，三折交叉验证）。五种解码器的性能在统计上没有显著差异：其平均 Dice 分数的跨度仅为 0.0071，低于大多数模型的折间变异；配对 Wilcoxon 符号秩检验显示，3D-EMCAD 与四种基线方法中的三种之间不存在显著差异。重要的是，3D-EMCAD 以最少的解码器参数（1.15 M）和运算量（8.6 G）实现了同等准确度，其运算量比 UNet++少 7.3×，而后者的得分并不更高。这些发现表明，对于体积脑肿瘤分割，超出适度预算的解码器容量并不能带来可测量的收益。**

索引词—**三维医学图像分割、脑肿瘤、BraTS、高效解码器、卷积注意力、深度卷积**

一、引言

胶质瘤是最常见且侵袭性最强的原发性脑肿瘤之一，其临床管理取决于
实现从多参数磁共振成像（MRI）中准确勾画肿瘤及其子区域。体积 MRI 扫描的手动勾画十分耗时，且容易受到较高评估者间变异的影响，因此推动了自动分割 [1], [2] 的发展。编码器—解码器网络已成为标准解决方案：编码器将体积压缩为层次化的语义特征，而解码器则从中重建出全分辨率分割结果 [3], [4]。

然而，架构创新在这两个组件之间的发展并不均衡。编码器经过了广泛的重新设计，而解码器则主要通过嵌套式跳跃路径、密集连接和注意力模块不断增加容量，这背后的假设是，更大的解码器必然能提高分割质量。对于体积分割而言，这一假设的代价尤其高昂，因为解码器需要在高空间分辨率下运行，密集 $k \times k \times k$ 卷积的成本随 k^3增长，而激活内存则随体积迅速增加。因此，解码器可能占用总计算预算中不成比例的份额，使典型临床环境中的硬件资源不堪重负。

一些成熟的解码器体现了这一向更大容量发展的趋势：UNet++ [5]引入了密集嵌套式跳跃路径，Attention U-Net [6] 通过学习得到的注意力系数对跳跃连接进行门控，而 SegResNet [7]采用了在脑肿瘤数据上极为有效的残差解码器；与此同时，自配置流水线 [8]和基于 transformer 的架构 [9], [10] 以更高的成本进一步提升了准确率。相反，CASCADE [11] 通过级联注意力解码器细化编码器特征，而 EMCAD [12]将多尺度深度卷积与通道、空间及分组门控注意力相结合，大幅降低了解码成本。然而，迄今为止，EMCAD 仅在二维场景中得到验证。恰恰是在内存和计算限制最为严峻的三维场景中，这种高效设计才能带来最大的收益。除这一架构空白之外，文献中还存在方法学空白：新的解码器通常与不同的编码器、训练计划一同引入，

通讯作者：hrkhan@eee.buet.ac.bd

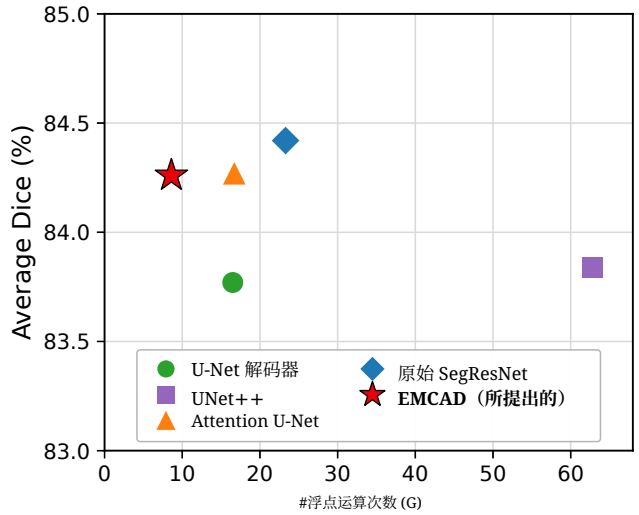

图 1：BraTS 2020 上平均 Dice 与解码器浮点运算次数的关系。3D-EMCAD 以最少的运算量达到其他解码器的精度；UNet++所需的浮点运算次数超过其七倍，却没有带来任何性能提升。

以及不同的数据划分。因此，性能提升无法明确归因于解码器本身，而且通常仅以汇总得分的形式报告，而未检验所观察到的差异是否具有统计学意义。

在本研究中，我们针对这两项不足开展工作。我们将 EMCAD 扩展至体积分割，并在受控实验方案下对其进行系统评估；在该方案中，解码器是唯一变量，评估基于 BraTS 2020 基准测试 [13]。具体而言，我们的贡献如下：(i) 提出 3D-EMCAD，即 EMCAD 解码器的一种体积形式，其中所有运算均为三维运算，且归一化方式针对体积图像训练所要求的小批量大小进行了调整；(ii) 对五种共享相同编码器、训练计划和数据折的解码器进行受控比较，并完整统计各解码器对应的参数和计算成本；(iii) 通过对全部 369 个病例进行配对显著性检验，表明五种解码器的精度在统计上无显著差异，而 3D-EMCAD 以最低的参数量和计算成本，达到规模大得多的解码器的性能。

II. 所提方法

本节介绍 3D-EMCAD，这是将面向二维分割的高效多尺度卷积注意力解码器（EMCAD） [12]扩展到体积图像后的方法。

A. 总体架构

该网络遵循图 2 所示的编码器—解码器范式。体积图像编码器将输入体积 $\mathbf{x} \in \mathbb{R}^{C_{\text{in}} \times H \times W \times D}$（包含 $C_{\text{in}} = 4$ 个 MRI模态）映射为一个四阶段特征金字塔 $\{X_1, X_2, X_3, X_4\}$，其阶段顺序由浅至深。解码器采用级联结构：首先细化最深层特征图，然后逐步上采样，并通过门控注意力机制与对应的跳跃连接融合，再进行细化。解码器生成四张分割图 $\{p_1, \ldots, p_4\}$，分别包含 $N = 3$ 个通道，每个阶段对应一张，且都恢复至输入分辨率；其中最精细的分割图 p_1 是最终预测，其余分割图则用于深度监督。

该设计与二维设计有三点不同：所有操作均在体积图像上进行；所有批归一化层均替换为组归一化（GN） [14]，，后者不依赖批次统计量，因此在体积图像训练所需的小批量大小下仍然稳定；此外，大核算子也针对三维进行了重新调整（第 II-C 节）。

B. 体积图像编码器

我们采用 3D SegResNet 骨干网络 [7]，，仅保留其收缩路径：先进行一次包含 $f = 16$ 个滤波器的初始卷积，随后经过五个残差阶段，各阶段分别包含 (1，2，2，4，4) 个块，阶段之间采用步幅为 2 的下采样。取最后四个阶段的输出，

$$X_i \in \mathbb{R}^{c_i \times \frac{H}{2^i} \times \frac{W}{2^i} \times \frac{D}{2^i}}, \qquad i = 1, 2, 3, 4, \tag{1}$$

输入的通道宽度为 $[32, 64, 128, 256]$，空间分辨率为 $1/2$、$1/4$、$1/8$ 和 $1/16$。此处评估的所有解码器均使用相同的编码器，以确保解码器是唯一的变量。

C. 多尺度卷积注意力模块（*MSCAM*）

MSCAM 通过依次应用通道注意力块（CAB）、空间注意力块（SAB）和多尺度卷积块（MSCB）来细化解码器的每个阶段。通道注意力块用于确定需要强调哪些通道，空间注意力块用于确定关注的位置，多尺度卷积块则用于增强不同感受野尺寸下的特征：

$$\text{MSCAM}(\mathbf{x}) = \text{MSCB}(\text{SAB}(\text{CAB}(\mathbf{x}))). \tag{2}$$

1) 通道注意力块：CAB 通过自适应最大池化 $P_m(\cdot)$ 和平均池化 $P_a(\cdot)$ 在整个体积上聚合每个通道，将两个描述符分别输入由逐点卷积 $C_1(\cdot)$、$C_2(\cdot)$ 构成的共享瓶颈，该瓶颈的压缩比为 $r = 16$，并包含一个中间 ReLU，最后通过 Sigmoid 函数 $\sigma(\cdot)$ 将两者之和转换为通道权重：

$$\begin{aligned} \mathbf{w} &= C_2\big(\text{ReLU}(C_1(P_m(\mathbf{x})))\big) \\ &\quad + C_2\big(\text{ReLU}(C_1(P_a(\mathbf{x})))\big), \\ \text{CAB}(\mathbf{x}) &= \sigma(\mathbf{w}) \circledast \mathbf{x}, \end{aligned} \tag{3}$$

其中，$\circledast$表示哈达玛积；在两条池化路径之间共享瓶颈层可降低该模块的成本。

2）空间注意力块：SAB 沿通道维度对最大值（$\text{Ch}_{\max}$）和平均值（Ch_{avg}）响应进行池化，将两个单通道体积进行拼接，并应用大卷积核卷积 LKC（·）和 Sigmoid 函数：

$$\text{SAB}(\mathbf{x}) = \sigma\big(\text{LKC}\big([\text{Ch}_{\max}(\mathbf{x}), \text{Ch}_{\text{avg}}(\mathbf{x})]\big)\big) \circledast \mathbf{x}. \tag{4}$$

由于体积卷积核的计算成本会随边长呈立方增长，而非平方增长，因此我们将 2D 设计中的 7×7卷积核替换为 $5 \times 5 \times 5$ 卷积核，在大幅降低成本的同时保留较大的局部上下文。

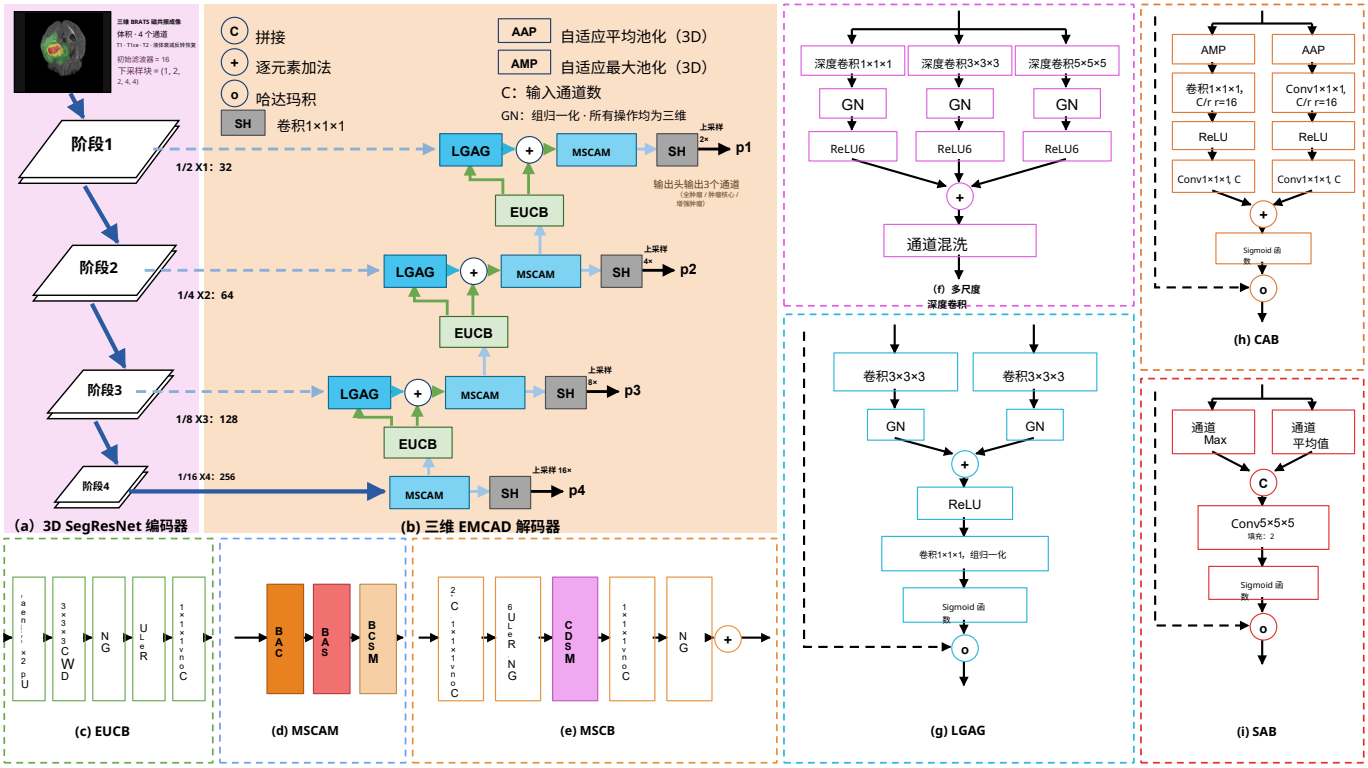

图 2：所提出的 3D-EMCAD 架构：(a) 生成金字塔 X_1–X_4 的体积 SegResNet 编码器；(b) 带有分割头 p_1–p_4 的级联解码器；(c) EUCB；(d) MSCAM；(e) MSCB；(f) 多尺度深度卷积（MSDC）；(g) LGAG；(h) CAB；以及 (i) SAB。所有运算均针对体积进行，且每个归一化层均采用组归一化（GN）。

3）多尺度卷积块：MSCB 遵循 MobileNetV2 [15]，的倒置残差设计，将单个深度卷积替换为多尺度深度卷积（MSDC）。它通过逐点卷积 $\mathrm{PWC}_1(\cdot)$、GN 和 ReLU6 将通道数扩展为两倍；应用 MSDC；通过通道混洗 $\mathrm{CS}(\cdot)$ [16] 恢复深度卷积滤波所破坏的通道间依赖关系；并通过 $\mathrm{PWC}_2(\cdot)$ 和 GN 将通道数投影回原值：

$$\begin{aligned} \mathbf{z} &= \mathrm{MSDC}\big(\mathrm{ReLU6}(\mathrm{GN}(\mathrm{PWC}_1(\mathbf{x})))\big), \\ \mathrm{MSCB}(\mathbf{x}) &= \mathrm{GN}\big(\mathrm{PWC}_2(\mathrm{CS}(\mathbf{z}))\big). \end{aligned} \tag{5}$$

MSDC采用一组具有不同卷积核尺寸的并行深度卷积，并对其响应求和，

$$\mathrm{MSDC}(\mathbf{x}) = \sum_{k\in\mathcal{K}} \mathrm{ReLU6}\big(\mathrm{GN}(\mathrm{DWC}_k(\mathbf{x}))\big), \tag{6}$$

其中，$\mathrm{DWC}_k(\cdot)$表示卷积核尺寸为$k\times k\times k$且具有 $\mathcal{K}=\{1,3,5\}$的深度卷积。这种并行结构能够在单个模块中捕获不同范围的病灶结构，同时深度卷积滤波降低了计算成本，这也是第IV节所报告的效率提升的主要来源。

D. 大核分组注意力门（*LGAG*）

跳跃连接传递高分辨率细节，但也会携带与目标无关的背景响应。LGAG通过使用由较粗尺度解码器特征导出的系数对跳跃特征进行门控，从而抑制这些背景响应。设$\mathbf{g}$为跳跃图，$\mathbf{u}$ 为上采样的解码器图，则二者均经由独立的 $3\times3\times3$ 卷积 $C_g(\cdot)$、 $C_u(\cdot)$变换、归一化、相加并激活：

$$q_{\mathrm{att}}(\mathbf{g},\mathbf{u}) = \mathrm{ReLU}\big(\mathrm{GN}(C_g(\mathbf{g})) + \mathrm{GN}(C_u(\mathbf{u}))\big), \tag{7}$$

$$\mathrm{LGAG}(\mathbf{g},\mathbf{u}) = \mathbf{g} \circledast \sigma\big(\mathrm{GN}(C(q_{\mathrm{att}}(\mathbf{g},\mathbf{u})))\big), \tag{8}$$

其中， $C(\cdot)$将门控压缩为单个通道。与 Attention U-Net [6]，中的逐点门控不同，式（7）中的 $3\times3\times3$卷积核能够基于更大的邻域估计门控，因此适用于胶质瘤亚区的弥散边界。

E. 高效上卷积块（*EUCB*）

EUCB通过三线性上采样Up(·)、带有GN和ReLU的 $3\times3\times3$深度卷积DWC(·)，以及逐点卷积$C_{1\times1\times1}(\cdot)$，将特征图的分辨率提高一倍，并使其通道数与下一个更浅层的阶段相匹配：

$$\mathrm{EUCB}(\mathbf{x}) = C_{1\times1\times1}\big(\mathrm{ReLU}(\mathrm{GN}(\mathrm{DWC}(\mathrm{Up}(\mathbf{x}))))\big). \tag{9}$$

深度方向滤波和插值，而非密集卷积和转置卷积，使该模块的计算成本显著低于传统上采样阶段。

F. 级联解码与分割头

解码从最深层开始，直接对 X_4 进行细化，然后逐步恢复至输入分辨率；对于 $i=$

3，2，1，将细化后的特征图$\mathbf{d}_{i+1}$上采样，并与其跳跃连接融合：

$$\mathbf{d}_4 = \mathrm{MSCAM}_4(X_4), \tag{10}$$

$$\mathbf{u}_i = \mathrm{EUCB}_{i+1}(\mathbf{d}_{i+1}), \tag{11}$$

$$\mathbf{d}_i = \mathrm{MSCAM}_i\big(\mathrm{LGAG}_i(X_i, \mathbf{u}_i) + \mathbf{u}_i\big). \tag{12}$$

(12) 的残差形式在门控抑制跳跃连接时保留上采样后的表示。每个阶段都以一个分割头 $\mathrm{SH}_i(\cdot)$ 结束，该分割头通过逐点卷积将 c_i 个通道映射到 N 个区域，随后通过三线性插值 $\mathcal{I}(\cdot)$ 将其调整至输入分辨率：

$$p_i = \mathcal{I}\big(\mathrm{SH}_i(\mathbf{d}_i)\big), \qquad i = 1, 2, 3, 4. \tag{13}$$

推理时，使用 Sigmoid 函数和 0.5 的阈值，从 p_1 得到三个嵌套区域。

G. 深度监督与训练目标

四个输出头均接受监督。对于预测值 p 和真实标注 y，我们结合软Dice项（用于应对体积肿瘤数据中严重的前景–背景不平衡）与二元交叉熵项（提供稳定的逐体素梯度）：

$$\mathcal{L}(p, y) = 1 - \frac{1}{N}\sum_{n=1}^{N} \frac{2\sum_v \hat{p}_{n,v}\, y_{n,v} + \epsilon}{\sum_v \hat{p}_{n,v} + \sum_v y_{n,v} + \epsilon} + \mathcal{L}_{\mathrm{BCE}}(p, y), \tag{14}$$

其中，$\hat{p} = \sigma(p)$，v 表示体素索引，n 表示 N 个区域，而 ϵ 用于避免除以零。总训练目标将四个输出头的损失相加，

$$\mathcal{L}_{\text{total}} = \sum_{i=1}^{4} w_i\, \mathcal{L}(p_i, y), \qquad w = (1, 0.5, 0.25, 0.125), \tag{15}$$

权重随深度按几何级数递减，因此，生成最终预测的输出头占主导地位，而较粗略的输出头则充当辅助正则化项。

III. 实验设置

A. 数据集与预处理

我们在 BraTS 2020 训练集 [1], [2], [13], 上评估所有模型。该数据集包含 369 例来自多家机构的术前胶质瘤病例，每例均提供四种配准并去除颅骨的 MRI 模态（T1、T1ce、T2 和 T2-FLAIR），分辨率为 $1\mathrm{mm}^3$ 各向同性。按照挑战赛协议，标签被转换为三个嵌套的评估区域：增强肿瘤（ET）、肿瘤核心（TC）和全肿瘤（WT）。由于 ET $\subseteq$ TC $\subseteq$ WT，这些区域以三个相互重叠的二值掩码表示，因此该任务属于多标签而非多类别任务。我们有意将预处理控制在最低限度，以确保比较反映的是解码器的差异，而非复杂的处理流程：每种模态均基于其非零体素归一化为零均值和单位方差，并将每个体积裁剪至其非零区域的边界框。

B. 基线解码器与受控方案

为单独考察解码器的贡献，我们将 3D-EMCAD 与四种成熟的替代方案进行比较：原生 SegResNet 解码器 [7]，（下文称为标准解码器）；体积 Attention U-Net 解码器 [6]；带有密集嵌套跳跃路径的体积 U-Net++ 解码器 [5]；以及标准 U-Net 解码器 [3], [4]。每种解码器均连接到与第 II-B 节相同的编码器，接收相同的特征金字塔，并使用相同的数据、数据折、训练计划和损失函数进行训练；五种配置之间唯一的区别是解码器。原生支持深度监督的解码器采用式（15）进行监督；其余解码器则采用式（14）的单头形式进行训练。

C. 实现细节

所有模型均使用 PyTorch 和 MONAI [17]实现，并在一块配备 16 GB 内存的 NVIDIA Tesla T4 GPU 上训练。训练时使用随机 $96 \times 96 \times 96$ 图像块，批量大小为 2；为应对严重的前景不平衡，以 0.7 的概率将图像块中心设为随机选取的全肿瘤体素，否则进行均匀采样。数据增强包括沿三个空间轴随机翻转，以及从 $[0.9, 1.1]$中抽取的乘性强度缩放。网络使用 AdamW [18] 优化 100 个周期，权重衰减为 10^{-5}；学习率在 8 个周期内线性预热至 10^{-4}，随后按照余弦调度衰减至 10^{-6}；训练全程将梯度裁剪至范数 0.5，并使用混合精度运算。推理时，采用滑动窗口推理对整个体积进行预测（窗口大小为 96^3，重叠率为 0.5，并使用高斯加权）；将最细粒度输出头的逻辑值通过 Sigmoid 函数处理，并以 0.5 为阈值进行二值化。此外，我们还报告了测试时增强（TTA）的结果，该结果通过对八种空间轴翻转组合下的概率图取平均获得。

D. 评估指标与统计分析 *is*

准确性采用 Dice相似系数（DSC）Y，$P = 2|Y \cap P|/(|Y| + |P|)$衡量，分别计算真实标注 Y 与预测结果 P在增强肿瘤、肿瘤核心和全肿瘤区域上的得分；我们报告三个区域评分及其平均值。所有结果均通过对 369 个病例进行三折交叉验证得出，数据折基于固定随机种子生成一次，并在每个解码器中重复使用；我们报告各折的平均值和标准差。效率通过仅针对解码器计算的参数量和浮点运算次数（FLOPs）来量化，基于单个 96^3 图像块。

由于准确率差异相较于折间变异较小，仅比较折均值无法确定某个解码器是否确实优于其他解码器。因此，我们对全部369个病例的逐病例平均Dice分数进行配对双侧Wilcoxon符号秩检验 [19]，将所提出的解码器与各基线进行比较；相较于配对 t检验，我们更倾向于采用Wilcoxon检验，因为逐病例Dice分数有界且呈明显偏态。采用Holm–Bonferroni方法 [20] 对由此得到的四个 p值进行多重检验校正，并以0.05为显著性水平进行评估。我们同时报告逐病例差值的中位数以及各 p值，因为在

如此大的样本量下，统计上可检测到的差异未必具有实际意义。

IV. 结果与讨论

A. 分割准确率

表 I 报告了五种解码器的性能，其最显著的特点是它们之间的差异极小。在不进行测试时增强的情况下，五种方法的平均 Dice 分数仅介于 0.8345 至 0.8416 之间，差距为 0.0071，小于大多数模型在不同数据折之间的标准差（0.0056 至 0.0145）：解码器造成的变化小于数据划分造成的变化。采用增强时情况也相同，平均值介于 0.8377 至 0.8442 之间。

3D-EMCAD 在不进行增强时取得最高的平均 Dice 分数（0.8416），而原始解码器在进行增强时取得最高分数（0.8442）；两者相差至多 0.0026。各区域的细分结果并未显示出始终占优的方法：最高 TC 分数来自原始解码器，不进行增强时最高 WT 分数来自 3D-EMCAD，而最高 ET 分数来自 Attention U-Net。增强带来了 0.0010 至 0.0034 的小幅一致提升，但并未改变排名；因此，以未增强结果作为主要比较依据。

B. 解码器效率

由于各解码器的准确率难分高下，成本便成为决定因素。图 1 和图 3 分别绘制了平均 Dice 与解码器浮点运算次数及参数量的关系；在两图中，所提出的解码器都位于准确率区间的最左端。

3D-EMCAD 需要 1.15 M 个参数和 8.6 G 次浮点运算，在这两项指标上都是五种方法中最小的。与 UNet++ 相比，差异尤为明显：后者需要 3.18 M 个参数和 62.8 G 次浮点运算，即参数量为 $2.8\times$ 、运算量为 $7.3\times$，但未经增强时的平均 Dice 低 0.0055。因此，其密集嵌套跳跃路径并未提升准确率，却大幅增加了成本。U-Net 和 Attention U-Net 解码器的成本介于这两个极端之间，约为前者的两倍，但准确率仍无优势；最接近的竞争者原始解码器虽然只需 1.22 M 个参数，却仍需要 23.3 G 次浮点运算，即 3D-EMCAD 的 $2.7\times$，因为它在全分辨率下使用密集卷积而非逐通道卷积。这种效率得益于式 (6) 中的多尺度深度卷积和式 (9) 中的逐通道上采样；在三维场景下，密集滤波的成本会随 k^3增长，从而进一步凸显这一优势。

C. 统计显著性

由于对折均值进行汇总比较无法确定这些微小差异是否有意义，因此我们采用第 III-D 节中的配对 Wilcoxon 符号秩检验在 369 个病例中的逐病例平均 Dice 分数。检验发现，3D-EMCAD 与 UNet++ （$p = 0.38$）、Attention U-Net （$p = 0.68$）或 U-Net（$p = 0.48$）解码器之间均无显著差异，所有 p 值均经过 Holm–Bonferroni 校正。只有与原始解码器（$p < 0.001$）相比时，差异才达到显著，其逐病例中位数优势为 0.0033 Dice。对于 369 个配对病例，该检验能够检测到这种幅度的差异，但三千分之一个

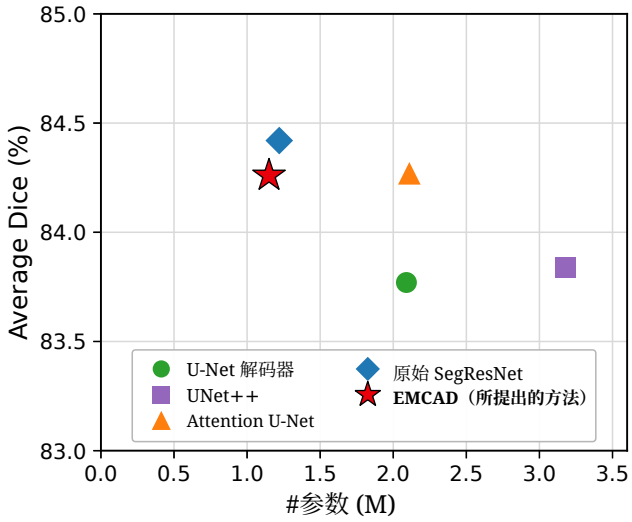

图 3：平均 Dice 与解码器参数量的关系。所提出的解码器在五种解码器中规模最小，同时保持在共同精度区间内。

Dice 点在实践中微不足道，远低于胶质瘤标注中报告的观察者间变异性 [1]。因此，这里的五种解码器在准确性上等效，恰当的表述应是同等性能下的效率，而非准确性更优。

D. 定性结果

图 4 展示了五种解码器在同一代表性验证病例的轴向切片上的预测结果以及真实标注。定量结果的一致性显而易见：每种解码器都能忠实地恢复全肿瘤的范围和核心区的位置；差异仅出现在边界处，尤其是薄薄的强化边缘；并且没有任何解码器表现出其他解码器未出现的失效模式。

E. 讨论与局限性

核心观察是，在固定编码器的情况下，解码器对准确率的影响非常小：从编码器金字塔重建分割所需的容量似乎并不高，且所考察的最小解码器已经能够满足这一需求。因此，在资源有限时，解码器容量是一个合理的缩减目标。

需要指出三点局限性。首先，评估仅使用了一个数据集；尽管 BraTS 2020 规模庞大且涵盖多个机构，但其对其他体积目标的泛化能力仍有待验证。其次，深度监督并未得到完全控制，因为 3D-EMCAD 和 UNet++采用多头结构训练，而其他模型则没有。鉴于准确率差异在统计上可以忽略不计，这不太可能改变结论，但这使得我们无法将剩余差异完全归因于架构本身。第三，效率方面的结论仅限于参数量和浮点运算次数：较低的浮点运算次数并不一定会按比例带来更低的延迟，因为与密集卷积相比，深度卷积和多分支操作在当前 GPU 内核上的映射效率较低。运行时行为的优化留待未来工作。

表 I：BraTS 2020（369 个病例，三折交叉验证）上的分割性能和解码器成本。区域评分为各数据折的均值；均值列还列出了各数据折间的标准差。#P 和 #F 分别表示解码器参数和浮点运算次数，不包括共享编码器，针对一个 96^3 图像块计算。各行按解码器大小排序，每列的最佳值以粗体显示；根据第 IV-C 节，除原始解码器外，这些差异均不显著，而原始解码器的差异在量级上可以忽略不计。

	解码器成本		Dice（无 TTA）				Dice（使用 TTA）			
解码器	#P (M)	#F(G)	TC	WT	ET	Mean	TC	WT	ET	Mean
3D-EMCAD（我们的方法）	**1.15**	**8.6**	0.8507	**0.9027**	0.7713	**0.8416±0.0098**	0.8537	0.9047	0.7693	0.8426±0.0100
原始 SegResNet	1.22	23.3	**0.8511**	0.9026	0.7687	0.8408±0.0145	**0.8547**	**0.9053**	0.7726	**0.8442±0.0146**
U-Net 解码器	2.09	16.5	0.8396	0.8996	0.7643	0.8345±0.0109	0.8446	0.9017	0.7668	0.8377±0.0097
Attention U-Net	2.11	16.7	0.8419	0.8999	**0.7778**	0.8399±0.0056	0.8473	0.9042	**0.7766**	0.8427±0.0044
UNet++	3.18	62.8	0.8477	0.8969	0.7638	0.8361±0.0067	0.8522	0.8992	0.7639	0.8384±0.0088

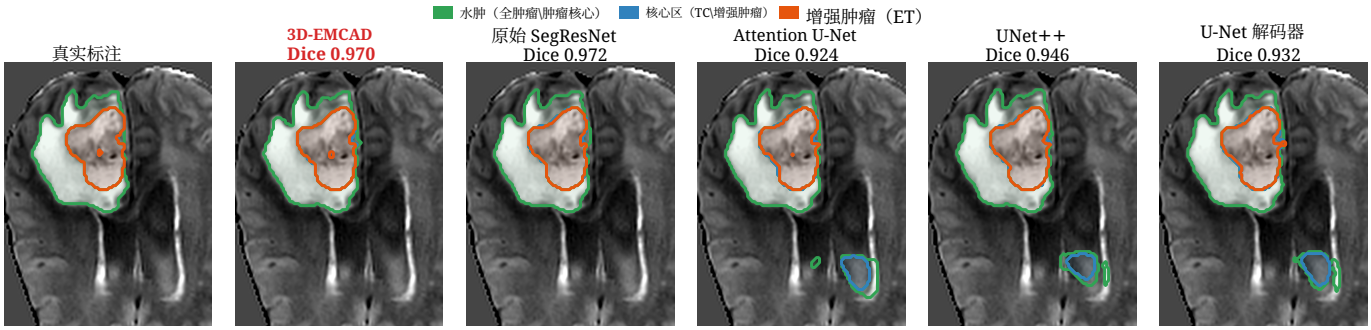

图4：代表性验证病例的定性比较；所有面板均显示相同的轴向切片，且裁剪区域和强度窗完全一致。

V. 结论

我们提出了 3D-EMCAD，即高效多尺度卷积注意力解码器的体积扩展，并在受控方案下将其与 BraTS 2020 上四种成熟的解码器进行了评估。在该方案中，编码器、训练计划和交叉验证折均保持固定，因此解码器是唯一变化的架构因素。

由此得出三项主要结论。首先，EMCAD 的效率优势可延续至体积数据场景：所提出的解码器仅需 1.15 M 个参数和 8.6 G 浮点运算次数，这两项指标均为五种解码器中最低，且其参数更少 2.8×，运算量更低 7.3×，相比 UNet++。其次，解码器复杂度的降低并未导致可测量的准确率损失：五种解码器的平均 Dice 差异最大为 0.0071，差异范围小于其中大多数解码器的折间标准差；对全部 369 个病例进行配对 Wilcoxon 符号秩检验后发现，3D-EMCAD 与 UNet++、Attention U-Net 或 U-Net 解码器之间均无显著差异。唯一显著的差异是相较于原始解码器，中位数差距仅为 0.0033，可忽略不计。第三，在此任务中，超过适度预算的解码器容量并未带来可测量的收益，因此在资源受限时，解码器是适合精简的部分。

因此，本文确立的主张是同等性能下的效率，具体针对参数量和运算量，而非延迟。未来的工作将考察这一结论是否适用于其他体积目标和编码器，并研究当前加速器上逐深度操作的运行时行为。

参考文献

[1] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby 等，“多模态脑肿瘤图像分割基准测试（BRATS）”，*IEEE*医学影像汇刊，第34卷，第10期，1993–2024页，2015年。[2] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby 等，“利用专家分割标注和影像组学特征推进癌症基因组图谱胶质瘤 MRI 数据集”，*Scientific Data*，第4卷，第170117页，2017年。[3] O. Ronneberger、P. Fischer 和 T. Brox，“U-Net：用于生物医学图像分割的卷积网络”，载于国际医学图像计算与计算机辅助干预会议（*MICCAI*）论文集，2015年，第234–241页。[4] ¨O. C¸ic¸ek、A. Abdulkadir、S. S. Lienkamp、T. Brox 和 O. Ronneberger，“3D U-Net：从稀疏标注中学习密集体积分割”，载于国际医学图像计算与计算机辅助干预会议（*MICCAI*）论文集，2016年，第424–432页。[5] Z. Zhou、M. M. Rahman Siddiquee、N. Tajbakhsh 和 J. Liang，“UNet++：一种用于医学图像分割的嵌套式 U-Net 架构”，载于深度学习医学图像分析、多模态学习与临床决策支持，2018年，第3–11页。[6] O. Oktay、J. Schlemper、L. Le Folgoc、M. Lee、M. Heinrich、K. Misawa、K. Mori、S. McDonagh、N. Y. Hammerla、B. Kainz、B. Glocker 和 D. Rueckert，“Attention U-Net：学习关注胰腺的部位”，*arXiv*预印本 *arXiv:1804.03999*，2018年。[7] A. Myronenko，“使用自编码器正则化进行 3D MRI 脑肿瘤分割”，载于国际 *MICCAI* 脑病变研讨会（*BrainLes*）论文集，2019年，第311–320页。[8] F. Isensee、P. F. Jaeger、S. A. A. Kohl、J. Petersen 和 K. H. Maier-Hein，“nnU-Net：一种用于基于深度学习的生物医学图像分割的自配置方法”，*Nature Methods*，第18卷，第2期，第203–211页，2021年。[9] A. Hatamizadeh、Y. Tang、V. Nath、D. Yang、A. Myronenko、B. Land-man、H. R. Roth 和 D. Xu，“UNETR：用于 3D 医学图像分割的变换器”，载于*IEEE/CVF*冬季计算机视觉应用会议（*WACV*）论文集，2022年，第574–584页。[10] A. Hatamizadeh、V. Nath、Y. Tang、D. Yang、H. R. Roth 和 D. Xu，“Swin UNETR：用于 MRI 图像中脑肿瘤语义分割的 Swin 变换器”，载于国际 *MICCAI* 脑病变研讨会（*BrainLes*）论文集，2021年，第272–284页。

[11] M. M. Rahman 和 R. Marculescu，“通过级联注意力解码进行医学图像分割”，载于*IEEE/CVF*冬季计算机视觉应用会议（*WACV*）论文集，2023年，第6222–6231页。[12] M. M. Rahman、M. Munir 和 R. Marculescu，“EMCAD：用于医学图像分割的高效多尺度卷积注意力解码”，载于*IEEE CVPR*论文集，2024年，第11 769–11 779页。[13] S. Bakas、M. Reyes、A. Jakab、S. Bauer、M. Rempfler、A. Crimi 等，“在 BRATS挑战赛中确定用于脑肿瘤分割、进展评估和总生存期预测的最佳机器学习算法”，*arXiv*预印本 *arXiv:1811.02629*，2018年。[14] Y. Wu 和 K. He，“组归一化”，载于欧洲计算机视觉会议（*ECCV*）论文集，2018年，第3–19页。[15] M. Sandler、A. Howard、M. Zhu、A. Zhmoginov 和 L.-C. Chen，“MobileNetV2：倒置残差与线性瓶颈”，载于*IEEE CVPR*，2018年，第 4510–4520 页。[16] X. Zhang、X. Zhou、M. Lin 和 J. Sun，《ShuffleNet：一种面向移动设备的极高效卷积神经网络》，载于 *IEEE CVPR* 会议论文集，2018年，第 6848–6856 页。[17] M. J. Cardoso、W. Li、R. Brown、N. Ma、E. Kerfoot、Y. Wang 等，《MONAI：面向医疗保健深度学习的开源框架》，*arXiv* 预印本 *arXiv:2211.02701*，2022年。[18] I. Loshchilov 和 F. Hutter，《解耦权重衰减正则化》，载于国际学习表征会议（*ICLR*）论文集，2019年。[19] F. Wilcoxon，《基于排序方法的个体比较》，《生物统计学通报》，第 1 卷，第 6 期，第 80–83 页，1945 年。[20] S. Holm，《一种简单的序贯拒绝式多重检验程序》，《斯堪的纳维亚统计学杂志》，第 6 卷，第 2 期，第 65–70 页，1979 年。